

## SEQ ID NO:1

TCNNATTCCCGAGAATAAATTTCTGTGACTAACTCTTCCTTTTGGTTGGTTCTTCATGGCATATGCCTTATGAAGG  
TAACATACCCAAGCTGCCTCTGCCTCCCGCAGTGAACCCCTACCTGCCCCTTTGGCAGGTTCTCTTACTGACCAT  
CCCCACCTGCCCCACACATCCTCCCCATGCACCCCAACTNTGAGCCCCCTCTGCTCAGTAAGTCTGTAGACTTG  
GTGGGTATATTGNNCTCATTGAGACTGCAGGCCCTTGGAGGGCAGGCTCTGACCTGCAGTAAGATGTGTGAGTGA  
TACTCAGCACACANTAGGTGGATAAATACCCCCACAGTAGGTGGGTAGTGAGCCCTGTGAGTCCACTGTAAGNCA  
CCATCTACATGGGCANAGCCTGCTTTAAGCGTGGGTAGGGACACAACAGTCTCTTCAGCAGGGCTTCTGGCACC  
ATCTACACAAGTCCATCCTCAGCTCTTCCACTCCCGGGTTCCCTCCTGGACCTGTGTGACTCTGAGGAACCTTGGG  
GAATTCCTAACCTCCCCCTTCAACTGAGCCCTTGGCTCTTGGAGTTAGCCACAACCTAACTACTCAGGTCCCTCC  
AACAAGGGGACTGTGTCTGTGGCTGGATGACTCATGCACACTGCTCCATCCCGCAATCTTGGCGGGACTTGGGC  
TGGGGAGGATGCCAGCCAGCTCAGGCTAGGAGCTTGCATCCTGTTGCCCCAACCCAGCCCTACCAGAACAGAGTG  
TACTCAGAGCTCCAGGACAAAAATCTGGAAACAGAGAGCCGGCTCTCATTTGGACCGAGATCTGAGTGATGAAAA  
GAGCAGGCAGAGGAAACAGCAAGTTCAAAGTTCCTGAGGTGGGAATGCGCTTGACACAACGAGACCTGAGAAGA  
ACACAGCAAAGGCCGTGTTACATTTGCTGNGACTCCAGCCCCAAGGATCTGGTCAGGACAGACATNGCGAGGA  
CTCACCTGGATAATCCAGAGCCATGGCCCATNACANGNNTNCTCTTTTTTTTTTTTCTTTCTTTTTCTTTTTCT  
TCTTTTTTTGNNNNNGGCCCCAAGACAGGCTTTCTTTGNGTAGCCCCGGCTGTTTTGGAACTNACTNTGTAGACC  
AAACTGGCCTGNGAACTCACAGAGATCCTCCTGNCCTTGNCTNCCGAGTACAAGGGTTAAAAGCCTGAGCCANTA  
CCACTGGCCAGGCTAACTAAGGTTCTTAACCTTTTTAAGNATTATTTTTCTTTCTTATGTATGTGTATATGGGGGA  
GGGGATGCACAAGGGCATGGGGGGGGGTCCCTGCAGAAGTCAGAAGAGGTGCCAGATCCCTGGGAGCTGGAATT  
AAAGTCAGTCATGAACATCCAAGATGGACACTGGGNACTGAACCTGGGTCCCTCTGCGAGAGGAGTAATGGTCT  
TAACTGCTGAGCCATCTCTAGGCCCAATGTCTGGTTTTGTTTTGTTTTGTTTTGTTTTGTTTTGTTTTGTTTTG  
TTTGTATTTGGGGGTTTTTGTGTTGCTGTTTGGTTGGTTTGTGTTTGGTTTTCTTGAGACAGGGTTTCTCTGTATA  
GCCCTGGCTGTCTGGAACCTACTCTATAGACTAGGCTGGCCTCGAACTCAGAAATCCTCCTGCCTCTGCCTCCC  
AAGTGCTGAGATTAAGGCCCGTGCCACCACTGCCCCGACGCCAATGTCTGTATTTTATTCTCTGTCAGAATCT  
CTTTTGTCTCCTAACGGAACATCATCCAGATTCTGGGAAGTACACTGAAGACAATGGGGTGGGTGTTGTTTCTC  
TCCTATGCCCTTTACATNCTCCCTACCTATTTAGATGTAACCATGATCTACCAGCTCATCACAGGCCACAGCTT  
AAACCTCCCTC

FIG. 1A

Alignment of sequences of two RapR71 and RapR72.

```

71      1  AGNAAGGAGAGGTAGGGTCAACACTGATTTCTGGCTTCAGAAATTCAGAGAATAAAATTT
72      1  -----TCN--NATTCCTGAGAATAAAATTT
consensus 1  .....** .. *****

71      61  CTGTGACTAACTCTTCCTTTTGTGGTTCTTCATGGNANCTGCCTTATGAAGGTAACATA
72      23  CTGTGACTAACTCTTCCTTTTGTGGTTCTTCATGGCATATGCCTTATGAAGGTAACATA
consensus 61  ***** * *****

71      121  CCCAAGCTGCCTCTGCCTCCCGCAGTGAACCCCTACCCCTGCCCTTTGGCAGGTTCTCTTA
72      83  CCCAAGCTGCCTCTGCCTCCCGCAGTGAACCCCTACCCCTGCCCTTTGGCAGGTTCTCTTA
consensus 121 *****

71      181  CTGACCATCCCCACCTGCCCCACACATCCTCCCCTATGCACCCCAACTCTGAGCCCCTCC
72      143  CTGACCATCCCCACCTGCCCCACACATCCTCCCCTATGCACCCCAACTNTGAGCCCCTCC
consensus 181 *****

71      241  TGCTCAGTAAGTCTGTAGACTTGGTGGGTATATTGGTCTCATTGAGACTGCAGGCCCTTG
72      203  TGCTCAGTAAGTCTGTAGACTTGGTGGGTATATTGGNCTCATTGAGACTGCAGGCCCTTG
consensus 241 *****

71      301  GAGGGCAGGCTCTGACCTGCAGTAAGATGTGTGAGTGATACTCAGCACACAGTAGGTGGA
72      263  GAGGGCAGGCTCTGACCTGCAGTAAGATGTGTGAGTGATACTCAGCACACANTAGGTGGA
consensus 301 *****

71      361  TAAATACCCCCACAGTAGGTGGGTAGTGAGCCCTGTGAGTCCACTGTAAGC-ACCATCTA
72      323  TAAATACCCCCACAGTAGGTGGGTAGTGAGCCCTGTGAGTCCACTGTAAGNCACCATCTA
consensus 361 *****

71      420  CATGGGCAGAGCCTGCTC-AAGCGTGGGTTANGGACACAACAGTTTCCTCAGAGGGCTTC
72      383  CATGGGCANAGCCTGCTTTAAGCGTGGGTTAGGGACACAACAGTTTNTTCAGAGGGCTTC
consensus 421 *****

71      479  TGGCACCATCCTACACAAGCCATNCTCAGCTCTTCCACTCCGGGTT-CCCTTCTGGACC
72      443  TGGCACCATN-TACACAAGCCATNCTCAGNTTTTCCACTNCCGGGTTTCCCTCCTGGACC
consensus 481 *****

71      538  TGTGTGACTCTGAGCAA-CTTGGGGAATTNCTAANCCTNCCCTTTCAACTGAAACCCCTTGG
72      502  TGTGTGACTCTGAGCNAACTTGGGGAATTCTTNAACCTTCCCTTTCAACTGNGCCCTTGG
consensus 541 *****

71      597  NTITGGANTTAA-CNNCAACCCTAACCTNCNNAAGNGCCCNCAACAANGGG-ACTGTG
72      562  NTCITGGNAATNNGGCAACCC-TAAC-TNCNCAAGNNCCCTCACAACAANGGGGACTGTG
consensus 601 ** **** * *****

```

FIG. 1B

## SEQ ID NO:2

ATGAGTGAAGTCTATACTCACAGGCACTGAGAAAGCCAGACTCAACGGCTA  
CCTCCTCCAAGATGTAACCATGATCTACCAGCTCATCACAGGCCACAGCT  
TAAACCTCCCTCCCTTTGTACATCTCCACCATCAACCACACCTTCCA  
TCTTTCTCTTTCATCTGACACATATCTTCCAACCTTCAGTCATCTAATAA  
GCAGACTTTAAAGCCACGGGTCTGGATATCCAATGGAAAATGACCAAA  
GGAAGAACACTTGCTCCTTAGTCCGACAAGAAGGTTTCAAAGGAGTCACC  
TTGCATGCTGAAGCACTTCCACAGAAGGAGCACCCCCCCCCCACCCTCA  
TCTGCAGGATTCCGAGATGGAAGAGAAGAGGCGAAAATATTCCATCAGCA  
GCGACAACCTGATACCACTGACGGTCACGTGACATCCACATCAGCATCA  
AGATGTTCCAAACTGCCAGCAGCACCAAGTCGGGCTGGCCCCGGCAGAA  
CGAGAAGAAGCCCTCAGAGGTTTCCGGACAGACTTGATCAGCCATGA  
AGATCCCAGATTCTACCAGCTCAGCCCGGATGACTACTACATCCTGGCG  
GACCCGTGGCGACAAGAATGGGAGAAAGGGGTGCAGGTACCTGCTGGAGC  
GGAGGCCATTCCAGAGCCTGTGGTGGGCTCCTCCACCCTGAAAGGCC  
CCCCACGCGAGATGTCCCCAGATAGCCCCACACTTGGTGAGGGTGCCAT  
CCTGACTGGCCAGGAGGCAGCCGCTACGACCTGGATGAGATCGATGCGTA  
CTGGTTGGAACCTCTCAACTCGGAGCTCAAGGAGATGGAGAAGCCCGAGC  
TGGATGAGCTTAACGTTAGAGCGTGTCTAGAGGAGCTAGAGACATTGTGC  
CACCAGAAATATGGCACAGGCCATTGAGACACAGGAGGGGCTGGGCATCGA  
GTACGACGAGGACGTTGTCTGCGACGTGTGCCGTTCCCCTGAAGGCGAGG  
ATGGCAACGAGATGGTCTTCTGTGACAAATGCAATGTCTGTGTGCACCAG  
GCATGCTACGGGATCCTCAAGGTGCCTACGGGCAGCTGGCTGTGCCGGAC  
CTGTGCCCTGGGAGTCCAGCCTAAGTGCCTGCTCTGCCCCAAGCGAGGAG  
GAGCCCTGAAGCCCACTAGAAGTGGGACCAAGTGGGTACACGTCACTGT  
GCCCTGTGGATTCTGAGGTGAGCATTGGCTGTCCAGAGAAGATGGAGCC  
CATTACCAAGATCTCGCATATTCCGGCCAGCCGCTGGGCCCTGTCTGCA  
GCCTCTGCAAGGAGTGACAGGTACCTGCATCCAGTGTTCATGCCTTCC  
TGCATCACAGCATTCCACGTTACGTGCGCCTTTGACCGAGGCCTGGAAAT  
GCGGACTATATTAGCTGACAAATGACGAGGTCAAGTTCAAGTCACTTTGCC  
AGGAGCACAGTGACGGGGGCCCTCGGAGTGAGCCTACTTCTGAGCCTGTG  
GAGCCAGCCAGGCCGTTGAGGATCTGGAAAAGGTGACCTTACGCAAGCA  
GCGGCTGCAGCAGCTGGAAGAAAACCTTCTATGAGCTAGTGGAGCCAGCTG  
AGGTGGCTGAACGGCTAGACCTGGCTGAGGCAGTGGTGGACTTCATCTAC  
CAGTACTGGAAGTTGAAGCGGAGAGCTAATGCCAACCAGCCGCTGTTGAC  
GCCCAAGACTGACGAGGTGGACAACCTGGCCCAACAGGAACAGGATGTCC  
TCTATCGACGCTGAAGCTTTTCAACCACCTGCGGCAGGACCTGGAGAGG  
GTAAGGAACCTGTGCTACATGGTGACAAGACGGGAGAGAACGAAACACAC  
CATCTGTAAACTTCAGGAGCAGATATTCCATCTACAGATGAACTTATTG  
AGCAAGACCTTTGAGAGAGCCTTCTGGGAGGAGGTCAAAGGGCAAGAAG  
AATGATTCAAAAAGGAAAGGCGGAGAGGTTCCAAAGGCGAGCAGCCCTGA  
GAAGAAAGAGAAAGTGAAGGCTGGGCCCCGAGTCTGTGCTGGGGCAGCTGG  
GTCTATCCACCTCGTTCCCCATCGACGGCACTTCTTCAACAGCTGGTTG  
GCACAGTCGGTTAGATCAGCAGAGGACATGGCCATGAGCGAGTGGTC  
TTTGAACAGTGGGCACCGGAGGATCCTGCTCCAGGTCTGCTGTGAGAGG  
AATTGCTACAAGATGAGGAGACGCTGCTCAGCTTCATGAGGGACCCCTCG  
CTACGACCTGGTGACCTTGCCAGAAAGGCCGAGGCCGACTCGCCTGCC  
TGCCAAAGAAACCATCCCCGCTGCAGGATGGGCCCCAGTGCACGGACCA  
CTCCAGACAAGCAACCCAAAGAAGGCCTGGGCCCAGGATGGCAAGGGGACG  
CAAGGACCACCCATGAGGAAGCCACCACGGAGGACGTCTTCTCATTTGCC  
GTCCAGCCCTGCAGCTGGGGACTGTCCAGTCCAGCAACACTGGAAAGCC  
CTCCACCACTGGCCTCCGAGATACTAGACAAGACAGCCCCCATGGCTTCC  
GACTTAAATGTCCAAGTGCTGGCCCTACAGTGAGCCCCAAACCTTGGG  
CAGGCTCCGGCCACCCGAGAGATGAAGGTCAAGTCCGAAATCTCCGGGTG  
CTAGATCCGATGTCTGGACAGGACTACCGTCTGCTGTGGCCGAGAGGCCA  
AAGGTCAGCCTGCATTTTGACACCGAGGCTGACGGCTACTTCTCTGATGA  
GGAGATGAGCGATTCTGAGGTAGAGGCAGAGGACAGTGGGGTACAACGAG  
CTTCCAGGGAGGCAGGGGCAGAGGAGGTGGTTCCGATGGGGTGCTGGCC  
TCCTAA

FIG. 2A

**SEQ ID NO:3**

MSESILTGTEKARLNGYLLQDVTMIYQLITGHSLNLPPLCHISTINHTLP  
SFSSSDTYLPTLQSSNKQTLKATGPGYPMENDQRKNTCSLVRQEGFKGVT  
LHAEALPTEGAPPPPHLQDSEMEEKRRKYSISSDNSDTTDGHVTSTAS  
RCSKLPSSSTKSGWPRQNEKKPSEVFRDITAMKIPDSYQLSPDDYYILA  
DPWRQEWKGVQVPAGAEAIPEPVVRLLPPLKGPPTQMSPDSPTLGEGAH  
PDWPGGSRDYLDDEIDAYWLELLNSELKEMEKPELDELTLERVLEELETLC  
HQNMAQAIETQEGLGIEYDEDVVCDCVCRSPEGEDGNEMVFCDKCNVCVHQ  
ACYGILKVPTGSWLCRTCALGVQPKCLLCPKRGALKPTRSGTKWVHVSC  
ALWIPEVSIQCPEKMEPITKISHIPASRWALSCSLCKECTGTCTIQCSPMS  
CITAFHVTCAFDRGLEMRITILADNDEVKFKSLCQEHSDGGPRSEPTSEPV  
EPSQAVEDLEKVTLRKQRLQQLLEENFYELVEPAEVAERLDLAEALVDFIY  
QYWKLKRRANANQPLLTPKTDEVNLAQQEQDVLRYRLKLFTHLRQDLER  
VRNLCYMVTRRERTKHTICKLQEQIFHLQMKLIEQDLCREPSGRRSKGKK  
NDSKRKGREGPKGSSPEKKEKVKAGPESVLGQLGLSTSFPIDGTFFNSWL  
AQSVQITAEDMAMSEWSLNSGHREDPAPGLLSEELLQDEETLLSFMRDPS  
LRPGDPARKARGRTRLPAKKKPSPLQDGPSARTTPDKQPKKAWAQDGKGT  
QGPPMRKPPRRTSSHLPSSPAAGDCPVPATLESPPPLASEILDKTAPMAS  
DLNVQVPGPTVSPKPLGRLRPPREMKVSRKSPGARSAGTGLPSAVAERP  
KVSLHFDTEADGYFSDEEMSDSEVEAEDSGVQRASREAGAEVVRMGVLAS

**SEQ ID NO:4**

MEEKRRKYSISSDNSDTTDGHVTSTASRCSKLPSSSTKSGWPRQNEKKPS  
EVFRDITAMKIPDSYQLSPDDYYILADPWWRQEWKGVQVPAGAEAIPE  
PVVRLLPPLKGPPTQMSPDSPTLGEGAHPDWPGGSRDYLDDEIDAYWLELL  
NSELKEMEKPELDELTLERVLEELETLC HQNMAQAIETQEGLGIEYDEDV  
VCDCVCRSPEGEDGNEMVFCDKCNVCVHQACYGILKVPTGSWLCRTCALGV  
QPKCLLCPKRGALKPTRSGTKWVHVSCALWIPEVSIQCPEKMEPITKIS  
HIPASRWALSCSLCKECTGTCTIQCSPMSCITAFHVTCAFDRGLEMRITILA  
DNDEVKFKSLCQEHSDGGPRSEPTSEPV EPSQAVEDLEKVTLRKQRLQQL  
EENFYELVEPAEVAERLDLAEALVDFIYQYWKLKRRANANQPLLTPKTDE  
VDNLAQQEQDVLRYRLKLFTHLRQDLERVRNLCYMVTRRERTKHTICKLQ  
EQIFHLQMKLIEQDLCREPSGRRSKGKKNDSKRKGREGPKGSSPEKKEKV  
KAGPESVLGQLGLSTSFPIDGTFFNSWLAQSVQITAEDMAMSEWSLNSGH  
REDPAPGLLSEELLQDEETLLSFMRDPSLRPGDPARKARGRTRLPAKKKPS  
PLQDGPSARTTPDKQPKKAWAQDGKGTQGPPMRKPPRRTSSHLPSSPA  
AGDCPVPATLESPPPLASEILDKTAPMASDLNVQVPGPTVSPKPLGRLRPP  
REMKVSRKSPGARSAGTGLPSAVAERP KVSLHFDTEADGYFSDEEMSDS  
EVEAEDSGVQRASREAGAEVVRMGVLAS

**FIG. 2B**

## SEQ ID NO:5

gttttaaaagaaacagaaacatacacaggggttggtgaatggtgccgaccgcggccatcgagttggaggctatttttgggggggatgga  
agagaagaggcgaaaatactccatcagcagtgacaactctgacaccactgacagtcagcgacatctacatccgcatcaagatgctccaaact  
gcccagcagcaccagtcgggtctggcccgacagaaagaaagccctcgaggttttccggacagacttgatcacagccatgaagatccc  
ggactcataccagctcagcccggtgactactacatcctggcagaccatggcgacaggaatgggagaaagggtgtgcaggtgcttgcggggc  
agaggccatcccagagccgtggtgaggatcctcccaccactggaaggcccccctgccagggcatccccagcagcaccatgcttggtagggg  
ctcccagcctgattggccagggggcagccgctatgacttggacgagattgatgcctactggctggagctcatcaactcggagcttaaggagat  
ggagaggccggagctggacgagctgacattagagcgtgtgtgctggaggagctggagaccctgtgccaccagaatatggccagggccattgagac  
gcaggaggggctgggcatcgagtacgacgaggatgtgtgtgctgcgacgtgtgtcgctcctcctgagggcgaggatggcaacgagatgggtcttctg  
tgacaagtgcacgtctgtgtgcatcaggcatgctacgggatcctcaagggtgccaccgggcagctggctgtgcggcagctgtgccctgggtgt  
ccagccaaagtgcctgtctgtgccccaaagcgaggaggagccttgaagcccaactagaagtgggaccaagtgggtgcatgtcagctgtgccctatg  
gattcctgaggtcagcatcggctgccagagaagatggagcccatcaccagatctcgcatatcccagccagccgctgggtctgtcctgag  
cctctgcaaggaatgcacaggcacctgcatccagtgctccatgccttctcctgcgtcacagcgttccatgtcacatgcgcctttgaccacggcct  
ggaaatgcggactatattagcagacaacgatgaggtcaagttcaagtcattctgccaggagcacagtgacgggggcccacgtaatgagccac  
atctgagcccaacgaaacccagccaggctggcgaggacctggaaaagggtgacccctgcgcaagcagcggctgcagcagctagaggaggacttcta  
cgagctgggtggagccggctgaggtggctgagcggctggacctggctgaggcactggctcgacttcatctaccagtaactggaagctgaaggagaa  
agccaatgccaaccagccgtgctgacccccaaagaccgacgaggtggacaacctggccagcaggagcaggacgtcctctaccgcccgtgaa  
gctcttcacccatctgcccagggacctagagaggggttagaaatctgtgctacatgggtgacaaggcgcgagagaaacgaaacacgcatctgcaa  
actccaggagcagatattccacctgcagatgaaacttattgaacaggatctgtgtcgagagcggctctgggaggagagcaagggaagaagag  
tgactcgaagaggaagggtgctgaggggtccaaaggcagcactgagaagaaagagaaagtgaaggcggggcctgactcagtcctggggcagct  
ggcaggcctgtccacctcattccccatcgatggcaccttcttcaacagctggctggcacagtcgggtgcagatcacagcagagaacatggccat  
gagcagtgaggccactgaacaatgggcaccgcgaggaccctgctccagggtgctgtcagaggaaactgctgcaggacgaggagacactgctcag  
cttcatgcccggaccctcgctgcgacctgggtgacctgctaggaaggcccgaggccgaccccgctgctgccaagaagaacaccaccacc  
accaccgcaggacgggctgggttccgggacgactccagacaaagcccccagaagacctggggccaggatgcaggcagtggaagggggtca  
agggccacctaccaggaagccaccacgtcggacatcttctcacttgcctgcagccctgcagccggggactgtcccatcctagccaccctga  
aagccccccgactggccctgagaccccgagaggcagcctcagtagctgtgactcagatgtccaagtgcctggccctgcagcaagccc  
taagcctttgggcccgtccggcccccgcgagagcaaggtaacccggagattgcccgggtgccaggcctgatgctgggatgggaccaccttc  
agctgtgggtgagaggcccaaggctcagcctgcattttgacactgagactgatggctacttctctgatggggagatgagcagactcagatgtaga  
ggccgaggacgggtgggtgacgggggtccccgggagggcaggggcagaggaggtggctccgcatgggctactggcctcctaactcaccacctt  
ccctgtcccaggccctgcccgtgggtccccccacaaggcctcagcccagtcacaactgccatttccagtcctctgctgagtgctccagaccctcga  
ggctgccactccgtcgtgggttttattttaatatagagagagttttgaattct

FIG. 3A

**SEQ ID NO:6**

MVPTAAIAVGGYFLGMEKKRRKYSISSDNSDTTDSHATSTSASRCSKLPSSSTKSGWPRQNEKKPSEVFRTDLIT  
 AMKIPDSYQLSPDDYYILADPWRQEWKGVQVPAGAEAIPEPVVRILPPLGPPAQASPSSTMLGEGSQPDWPGG  
 SRYDLDEIDAYWLELINSELKEMERPELDELTLERVLEELETCHQNMAIAIETQEGLGIEYDEDDVCDVCRSPE  
 GEDGNEMVFCDKCNVCVHQACYGILKVPTGSQLCRTCALGVQPKCLLCPKRGGALKPTRSGTKWVHVSCALWIPE  
 VSIGCPEKMEPITKISHIPASRWALSCSLCKECTGTCTIQCSMPSCVTAHFVTCAFDHGLEMRTILADNDEVKFKS  
 FCQEHSDGGPRNEPTSEPTSPQAGEDLEKVTLRKQRLQLEEDFYELVEPAEVAERLDLAEALVDFIYQYWKLK  
 RKANANQPLLTPTKTEVDNLAQQEQDVLRYRLKLFTHLRQDLERVRNLCYMTTRRERTKHAICKLQEQIFHLQMK  
 LIEQDLCRERSGRRAGKKSDDSKRKGCEGSKGSTEKKEKVKAGPDSVLGQLAGLSTSFPIDGTFFNSWLAQSVQI  
 TAENMAMSEWPLNNGHREDPAPGLLSEELLQDEETLLSFMRDPSLRPGDPARKARGRTRLPAKKKPPPPPPQDGP  
 GSRTTPDKAPKKTWGQDAGSGKGGQGPTRKPPRTSSHLPSSPAAGDCPILATPESPPPLAPETPDEAASVAAD  
 SDVQVPGPAASPKPLGRLRPPRESKVTRRLPGARPDAGMGPPSAVAERPVSLLHFDTTETDGYFSDGEMSDSDVEA  
 EDGGVQRGPREAGAEVVRMGVLAS

**SEQ ID NO:7**

MEEKRRKYSISSDNSDTTDSHATSTSASRCSKLPSSSTKSGWPRQNEKKPSEVFRTDLITAMKIPDSYQLSPDDYY  
 ILADPWRQEWKGVQVPAGAEAIPEPVVRILPPLGPPAQASPSSTMLGEGSQPDWPGGSRYDLDEIDAYWLELI  
 NSELKEMERPELDELTLERVLEELETCHQNMAIAIETQEGLGIEYDEDDVCDVCRSPEGEDGNEMVFCDKCNVC  
 VHQACYGILKVPTGSQLCRTCALGVQPKCLLCPKRGGALKPTRSGTKWVHVSCALWIPEVSIGCPEKMEPITKIS  
 HIPASRWALSCSLCKECTGTCTIQCSMPSCVTAHFVTCAFDHGLEMRTILADNDEVKFKSFCQEHSDGGPRNEPTS  
 EPTSPQAGEDLEKVTLRKQRLQLEEDFYELVEPAEVAERLDLAEALVDFIYQYWKLK RKANANQPLLTPTKTE  
 VDNLAQQEQDVLRYRLKLFTHLRQDLERVRNLCYMTTRRERTKHAICKLQEQIFHLQMKLIEQDLCRERSGRRAK  
 GKKSDSKRKGCEGSKGSTEKKEKVKAGPDSVLGQLAGLSTSFPIDGTFFNSWLAQSVQITAENMAMSEWPLNNGH  
 REDPAPGLLSEELLQDEETLLSFMRDPSLRPGDPARKARGRTRLPAKKKPPPPPPQDGPGRSRTTPDKAPKKTWGQ  
 DAGSGKGGQGPTRKPPRTSSHLPSSPAAGDCPILATPESPPPLAPETPDEAASVAADSDVQVPGPAASPKPLG  
 RLRPPRESKVTRRLPGARPDAGMGPPSAVAERPVSLLHFDTTETDGYFSDGEMSDSDVEAEDGGVQRGPREAGAE  
 VVRMGVLAS

**FIG. 3B**

## Exon 1 (SEQ ID NO:8)

GGGGGTGGGTGAATGGTGCCGACCGCGGCCATCGCAGTTGGAGGCTATTTTTTGGGGGGGGTGAGTAGCGTCCAT  
GGAGTTACTTTGCGCCCACTCCTAGCGGCACCGGCTTAGGTCTGCGGGCCGACCGTCCCCGGCGGGGGCGTGG  
GGCCTGGGACGCGCGGGCCCCGGCCGCTCCTCGCCGCGACCCCGGATGGATGCGCGCCCCCGCCCTCCCGCG  
CCGGCCCCAGGAGCTCCCGGCTTCGGGAGCATCCTTCCCGCGCCGGTCCCTGCAGCGGCGCGTAGCCGAGGGCAG  
CGCCCGTCAGGGGGGCACCGCGGAGCAAG

## Exon 2 (SEQ ID NO:9)

ATGGAAGAGAAGAGGCGAAAATACTCCATCAGCAGTGACAACTCTGACACCACTGACA

## Exon 3 (SEQ ID NO:10)

GTCATGCGACATCTACATCCGCATCAAGATGCTCCAACTGCCAGCAGCACCAAGTCGGGCTGGCCCCGACAGA  
ACGAAAAGAAGCCCTCCGAG

## Exon 4 (SEQ ID NO:11)

GTTTTCCGGACAGACTTGATCACAGCCATGAAGATCCCGGACTCATACCAGCTCAGCCCGGATGACTACTACATC  
CTGGCAGACCCATGGCGACAGGAATGGGAGAAAGGTGTGCAGGTGCCTGCCGGGGCAGAGGCCATCCCAGAGCCC  
GTGGTGAG

## Exon 5 (SEQ ID NO:12)

GATCCTCCCACCACTGGAAGGCCCCCTGCCAGGCATCCCCGAGCAGCACCATGCTTGGTGAGGGCTCCCAGCC  
TGATTGGCCAGGGGGCAGCCGCTATGACTTGACGAGATTGATGCCTACTGGCTGGAGCTCATCAACTCGGAGCT  
TAAGGAGATGG

## Exon 6 (SEQ ID NO:13)

AGAGGCCGGAGCTGGACGAGCTGACATTAGAGCGTGTGCTGGAGGAGCTGGAGACCCTGTGCCACCAGAATATGG  
CCAGGGCCATTGAGACGCGAGGAGGGGCTGGGCATCGAGTACGACGAGGATGTTGTCTGCGACGTGTGTCTGCTCTC  
CTGAGGGCGAGGATGGCAACGAGATGGTCTTCTGTGACAAGTGCAACGTCTGTGTGCATCAG

## Exon 7 (SEQ ID NO:14)

GCATGTACGGGATCCTCAAGGTGCCACGGGCAGCTGGCTGTGCCGGACGTGTGCCCTGGGTGTCCAGCCAAAG  
TGCCTGTCTGCCCCAAGCGAGGAGGAGCCTTGAAGCCCACTAGAAGTGGGACCAAGTGGGTGCATGTCAGCTGT  
GCCCTATGGATTCTCTGAG

## Exon 8 (SEQ ID NO:15)

GTCAGCATCGGCTGCCAGAGAGAAGATGGAGCCCATACCAAGATCTCGCATATCCCAGCCAGCCGCTGGGCTCTG  
TCCTGCAGCCTCTGCAAGGAATGCACAGGCACCTGCATCCAG

## Exon 9 (SEQ ID NO:16)

TGTTCCATGCCTTCTGCGTCACAGCGTTCCATGTACATGCGCCTTTGACCACGGCCTGGAAATGCGGACTATA  
TTAGCAGACAACGATGAGGTCAAGTTCAAGTCATTCTGCCAGGAGCACAGTGACGGGGGCCACGTAATGAGCCC  
ACATCTGAGCCACGGAACCCAGCCAGGCTGGCGAGGACCTGGAAAAGGTGACCCCTGCGCAAGCAGCGGCTGCAG  
CAGCTAGAGGAGGACTTCTACGAGCTGGTGGAGCCGGCTGAGGTGGCTGAGCGGCTGGACCTGGCTGAGGCACTG  
GTCGACTTCATCTACCACTACTGGAAGCTGAAGAGGAAAGCCAATGCCAACCAGCCGCTGCTGACCCCAAGACC  
GACGAGGTGGACAACCTGGCCCAGCAGGAGCAGGACGTCCTTACCGCCGCTGAAGCTCTTACCCATCTGCGG  
CAGGACCTAGAGAGG

## Exon 10 (SEQ ID NO:17)

GTTAGAAATCTGTGCTACATGGTGACAAGGCGGAGAGAACGAAACAGCCATCTGCAAACTCCAGGAGCAGATA  
TTCCACCTGCAGATGAACTTATTGAACAGGATCTGTGTGAG

FIG. 3C

## Exon 11 (SEQ ID NO:18)

GCCTGTCCACCTCATTCCCCATCGATGGCACCTTCTTCAACAGCTGGCTGGCACAGTCGGTGCAGATCACAGCAG  
AGAACATGGCCATGAGCGAGTGGCCACTGAACAATGGGCACCGCGAGGACCCTGCTCCAGGGCTGCTGTCAGAGG  
AACTGCTGCAGGACGAGGAGACACTGCTCAGCTTCATGCGGGACCCCTCGCTGCGACCTGGTGACCCTGCTAGGA  
AGGCCCCGAGGCCGACCCGCCCTGCCTGCCAAGAAGAAACCACCACCACCACCACCGCAGGACGGGCCCTGGTTCAC  
GGACGACTCCAGACAAAGCCCCCAAGAAGACCTGGGGCCAGGATGCAGGCAGTGGCAAGGGGGGTCAAGGGCCAC  
CTACCAGGAAGCCACCACGTCGGACATCTTCTCACTTGCCGTCCAGCCCTGCAGCCGGGGACTGTCCCATCCTAG  
CCACCCCTGAAAGCCCCCGCCACTGGCCCCCTGAGACCCCGGACGAGGCAGCCTCAGTAGCTGCTGACTCAGATG  
TCCAAGTGCCTGGCCCTGCAGCAAGCCCTAAGCCTTTGGGCCGGCTCCGGCCACCCCGCAGAGCAAGGTAACCC  
GGAGATTGCCGGGTGCCAGGCCTGATGCTGGGATGGGACCACCTTCAGCTGTGGCTGAGAGGCCCAAGGTCAGCC  
TGCATTTTGACACTGAGACTGATGGCTACTTCTCTGATGGGGAGATGAGCGACTCAGATGTAGAGGCCGAGGACG  
GTGGGGTGCAGCGGGGTCCCCGGGAGGCAGGGGCAGAGGAGGTGGTCCGCATGGGCGTACTGGCCTCC

FIG. 3C (continued)



SEQ ID NO: 19

**Promoter and regulatory region:**

cgccctggggacagggcgggctagggcgccccagagtcctatggggagtcggggccaggggtgccagcaggcgt  
gggtggtggggctgcgagggagggcacccttccccacggggcccgcaacgctacctggactccccgcggagcca  
aacaactgggcggggggttggggggcgcgacgggggtgtcgggagcggagatccgagtgaataagaaaaaagt  
ggctactccccctccctcgtcctcctgccccccccacccccacccccacccaacacatttttttttctaaag  
agatcacaaaggaagtcttgggttaaaaaagaaacagaaacatacacaggggggttgggtgaatgggtgccgaccgcggc  
catcgagttggaggctatttttggggggggtgagtagcgtccatggagttactttgcgcccactcctagcggc  
accggcttaggtcctgcgggcccagcgtccccggcgggggcggtggggcctgggacgcgcgggccccggcgcct  
ccctcgccgcgaccccggtggatgcgcgcccccccgccctcccgcgccggccccaggagctcccggttcgggag  
catccttcccgcgccggtccctgcagcggcggtagccgagggcagcgcccggtcaggggggacccgcggagcaag  
gtaagatccagcccccgcggtgggcccctgcgcacatctccacgagcttatttggcggtttttgcaacagatctgcc  
agcgtccttcgctccctcgtcctccttgcctcgtcgtccctcctcctcctgctgggtgcctgttctaggaagc  
cagcgcgagaggggggggatgcacagcacaggggagagagattgcgcagttggtcagtcgtgttttaaagagt  
acagtgcggggaggctgagaggggcgcagcaacaacaacttttggagggtgagcttggcgaccttctttatta  
atgactgcggcaaagcgcggcgggcgagggggcgcgggcgggcgggcgccagggtgcaacttccc  
cgcggtcccgccggcgtaggggctgcggcgggagatgggtacgggtggggaggtcgagcgcccgggcggggg  
ctccgagaacctggagctatctgcctcctgtctccccgagtttcattttgttgatacgagcacgtccgggcgc  
cgaaccgggctgagccggtgcacatgacctgcgctgggtcacgtgcagccggtccggtcccagacaccttccg  
ggggccaccgcctccgcccgtgcgccccctctcccgcccggtgcacgcgggcgtgcacgcggggcagcatgc  
tcggctcctgggggtggaggctctgcacaaattagacagtttttttggaggggcgggggacacccttccagggtg  
agtgtggagggtgcg

**FIG. 3D**

# SEQ ID NO: 19

## Promoter 2.0 Prediction Results

cgccctcggggacagggcgggctagggcgccccagagtcacatggggagtcggggccag  
 ggtgccagcagggcggtgggtggggctgcgagggagggcacccttccccacggggcccg  
 caacgctacgtggactccccgcccggagccaaacaactggcggggggttggggggcggc  
 gacgggggtgtcgggagcggagatccgagtgaaataagaaaaagtggctactccccctcc  
 ctgctcctcctgcccccccccccccccccccccaacacattttttttctaaag  
 agatcacaaaggaagtccttggtttaaaaagaaacagaaacatacacagggggttggtgaat  
 ggtgcgcagccgcccacatcgagtgagggtatttttgggggggtgagtagcgcca  
 tggagttactttcgcccactcctagcgccacgggttaggtcctgcccggccgaccgtcc  
 ccggcgggggcggtggggcctgggacgcgcggggcccgccgctccctcgccgcgaccc  
 cggatggatgcgcgcccccgccctcccgcgccggccccaggagctcccggttcgggag  
 catccttcccgccgggtccctgcagcgccgctagccgagggcagcgccgtcaggggg  
 gcaccgaggagcaaggaagatccagccccggcgagtgggccctgcgcacacacgac  
 gtatttggcggttttgaacagatctgcagcgctcttcgctccctcgctctctctgc  
 tcgctcgctccctctctctctctgctggtgctgttctaggaagccagcgcgagagggg  
 ggggatgcacagcacaggggagagagattgcgcagtggtcagtcgtgttttaagagt  
 acagtcggggagggctgagagggcgcatgcaacaacaacttttgaagggtagcttgg  
 cgaccttcttattaatgactgcggcaaaagcggccccggcgcgagggggcgcgggcg  
 ggcgggggcgcgccagggtgcaacttcccccggggtccggcgggcgtaggggctgcg  
 gcgggagatgggtacggtggggaggtcgagcgcccgggcgggggctccgagaacctgga  
 gctatctgcctcctgtctccccagagtttcattttgttgatacgcagcagtcggggcg  
 cgaacggggtgagcggggtgcacatgacctcgcgctggggtcacgtgcagcggtccggg  
 ccagacaccttccgggggcccacgcctccgcctgtcgccctctccggggcggtgc  
 acggggcgctgcagcgggggcagcatgctcggtcctggggttgagggtctgcacaa  
 attagacagtttttttggagggcgggggacaccctttccaggtgagtgaggaggtgcg

## PREDICTED TRANSCRIPTION START SITES:

Sequence, 1440 nucleotides

Position Score Likelihood

500 1.072 Highly likely prediction  
 1100 0.587 Marginal prediction

Promoter predictions for 1 eukaryotic sequence with score cutoff 0.80:

Promoter predictions:

Start	End	Score	Promoter Sequence	138	188	0.80
CCCGCCGGAGCCAAACA	ACTGGGCGGGGGCTT	GGGGGGGGCGGCACGGGG	(SEQ ID NO:20)	481	531	0.88
CCGGCGGGGGCGGTGGG	CGCTGGGACGCGCGGG	CCGGCGCGCTCCCTC	(SEQ ID NO:21)	963	1013	0.98
ACCTCTTTTATTAATG	ACTGCGGCAAGCGCCC	CGGGCGGCGAGGGGG	(SEQ ID NO:22)	992	1042	0.84
GCCCCCGGGCGGCGAG	GGGGCGCGGGCGGGG	CGGCCAGGGCTG	(SEQ ID NO:23)			

FIG. 3E

**FIG. 4A**

[illegible]

**FIG. 4A (continued)**

M 1 WSESLTCTEKARLNGYLLQDVMTIYQLITGHSLNLPPLCHITSTIMETLPSFSSSETYLPTLQSSNKOTLKATGPGYPMENDQRKH"CSI  
H 1 .....  
consensus 1 .....  
  
M 91 VRQEGKGVTLHAEALPTEGAPPPPPHLDSEMEKKRRKYSISSDSDITIDCHMTSASRCSKLPSSSTKSGWPRONEKKPSEVFR"DLI  
H 1 -----MVPTAAIAVGGYFLQGMEKKRRKYSISSDSDITIDCHMTSASRCSKLPSSSTKSGWPRONEKKPSEVFR"DLI  
consensus 91 ..... \* .....  
  
M 181 TAMKIPDSYQLSPEDYYITLADPNRQWFRGVOVPAGAAEAIPEPVVPLPPLKSPPTTNSPSTPIIGECALPDKPGSSRYDLEEDAYWLE  
H 75 TAMKIPDSYQLSPEDYYITLADPNRQWFRGVOVPAGAAEAIPEPVVPLPPLKSPPTTNSPSTPIIGECALPDKPGSSRYDLEEDAYWLE  
consensus 181 ..... \* .....  
  
M 271 LKNSFLKEMERPELDELTLERVLEELTLCQNMAAIITQDGLGIEYDEYVVDVCRSPGEGLEGNHWFCDRKNVCVHQACYGILKVP  
H 165 LKNSFLKEMERPELDELTLERVLEELTLCQNMAAIITQDGLGIEYDEYVVDVCRSPGEGLEGNHWFCDRKNVCVHQACYGILKVP  
consensus 271 \* .....  
  
M 361 GSWLCRTCALGVQPKCLCPFRGGALKPTRSGTKVHVIVSCALNIPEVSIQPEKMEPIIKISHIPASRWALSCSLCKECTGICICQCSNPS  
H 255 GSWLCRTCALGVQPKCLCPFRGGALKPTRSGTKVHVIVSCALNIPEVSIQPEKMEPIIKISHIPASRWALSCSLCKECTGICICQCSNPS  
consensus 361 .....  
  
M 451 CHTAFHVITCAFDHGLEFMRITLADNDEVKFKSLCOEHSDCGPPSPPTSEPVSPSOAVEDLEKVTIRKQRQQIFNFYELVEFAEVAERLL  
H 345 CHTAFHVITCAFDHGLEFMRITLADNDEVKFKSLCOEHSDCGPPSPPTSEPVSPSOAVEDLEKVTIRKQRQQIFNFYELVEFAEVAERLL  
consensus 451 \* .....  
  
M 541 LAEALVDFIYQYWKLPKXANANQPLLTPTKTEVDNLAQEQDVLVLRPLKLFTHLRQDLERVMICVMVTRERTIKHAIICKLCEQIFHLQY  
H 435 LAEALVDFIYQYWKLPKXANANQPLLTPTKTEVDNLAQEQDVLVLRPLKLFTHLRQDLERVMICVMVTRERTIKHAIICKLCEQIFHLQY  
consensus 541 .....  
  
M 631 KLIEQDLCEPSCRRSNGKKNDSKRKREDCPKGSSPEKKEVKVAGPSEVLSQLSLSTSPFPIDETFFNWSLACSVQITALVWASENPLN  
H 525 KLIEQDLCEPSCRRSNGKKNDSKRKREDCPKGSSPEKKEVKVAGPSEVLSQLSLSTSPFPIDETFFNWSLACSVQITALVWASENPLN  
consensus 631 .....  
  
M 720 SHREDPAPGLLSEILLQDEETLLSFMWRDPSRFQDPAPKAKTRTRIPAKKKPSF-----LDQGPSARTTDPKAPKKAQODE---KDTGPP  
H 614 SHREDPAPGLLSEILLQDEETLLSFMWRDPSRFQDPAPKAKGRTPLPAKKKPPPEPFDGPGSRTTDPKAPKKAQODE---KDTGPP  
consensus 721 .....  
  
M 805 MRKPPRRSSSLPSSPAAGDCHVPAATLESPPELASILLKTNPPASLNVQVPGITVSPKPLGRLLRPREKVKRKSQALSDACTSLPS  
H 704 TRKPPRRSSSLPSSPAAGDCHVPAATLESPPELASILLKTNPPASLNVQVPGITVSPKPLGRLLRPREKVKRKSQALSDACTSLPS  
consensus 811 .....  
  
M 895 AVAEKPKVSLHFDTEADGYFSLCEMSDSHVAEDSIVQVPSREAGAEVVRMGVLAS  
H 794 AVAEKPKVSLHFDTEADGYFSLCEMSDSHVAEDSIVQVPSREAGAEVVRMGVLAS  
consensus 901 .....

**FIG. 4B**

10/524426

# UCSC Genome Browser on Mouse Feb. 2002 Freeze

move <<< << < > >> zoom in 1.5x 3x 10x zoom out 1.5x 3x 10x  
position chr11:52214824-52353638 size 138815, image width 1500 jump

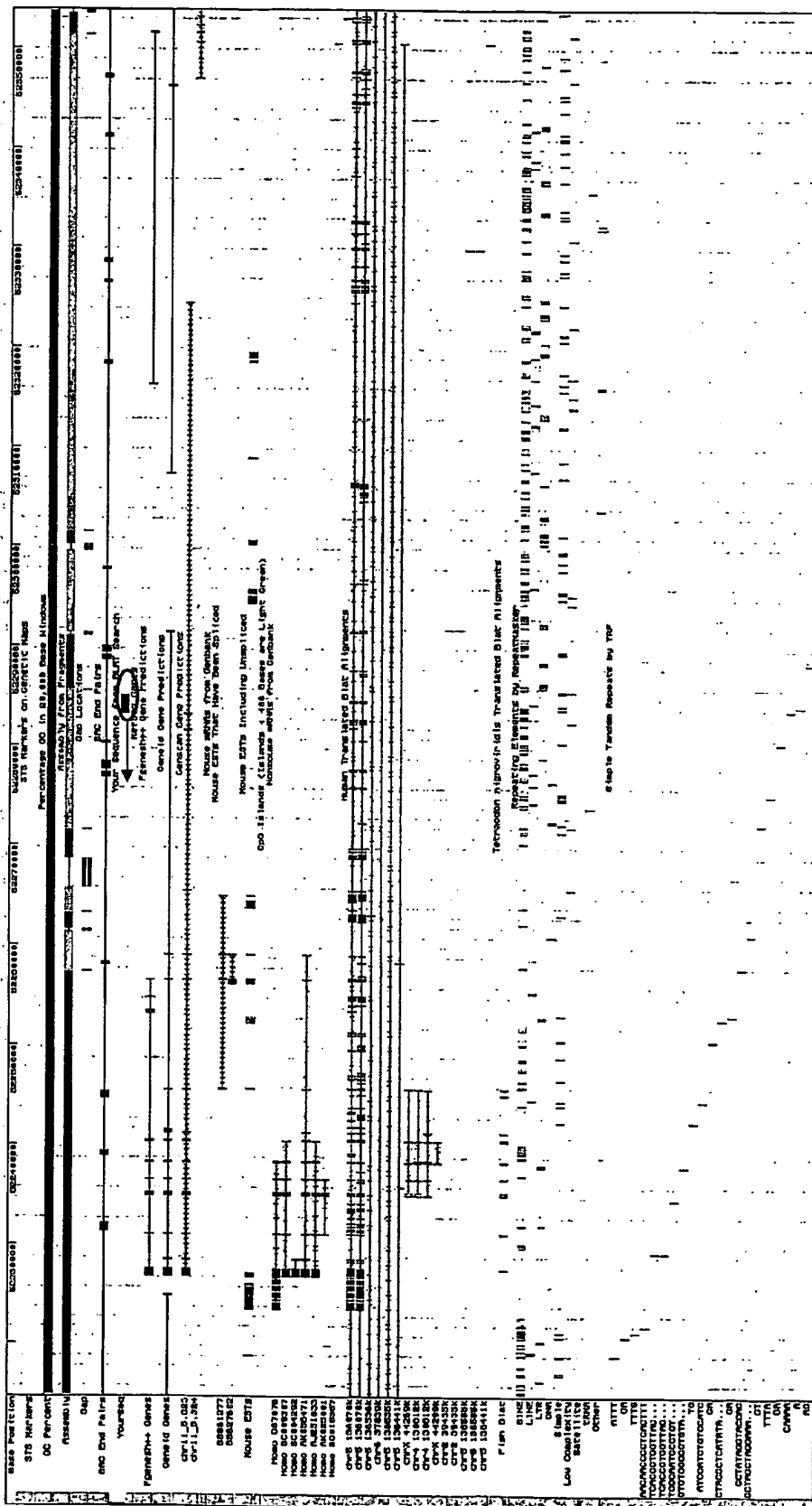


FIG. 5A

RapR7

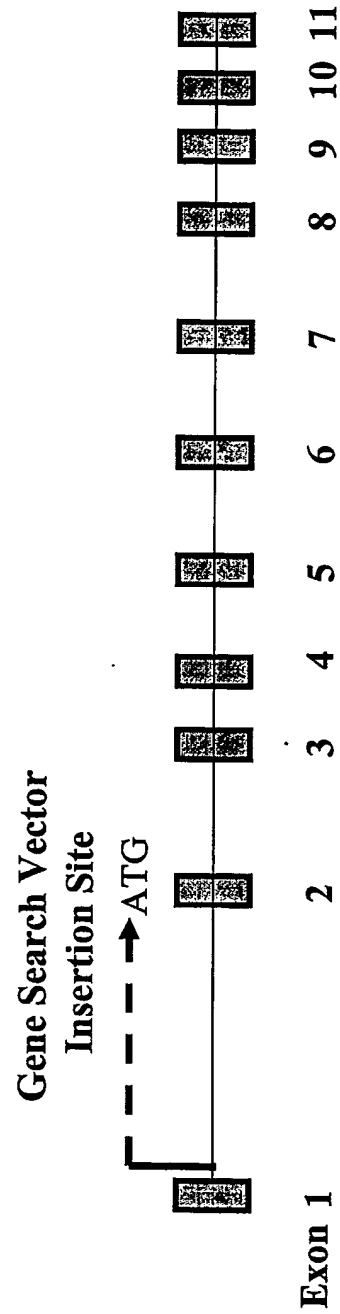


FIG. 5B



**FIG. 6A**

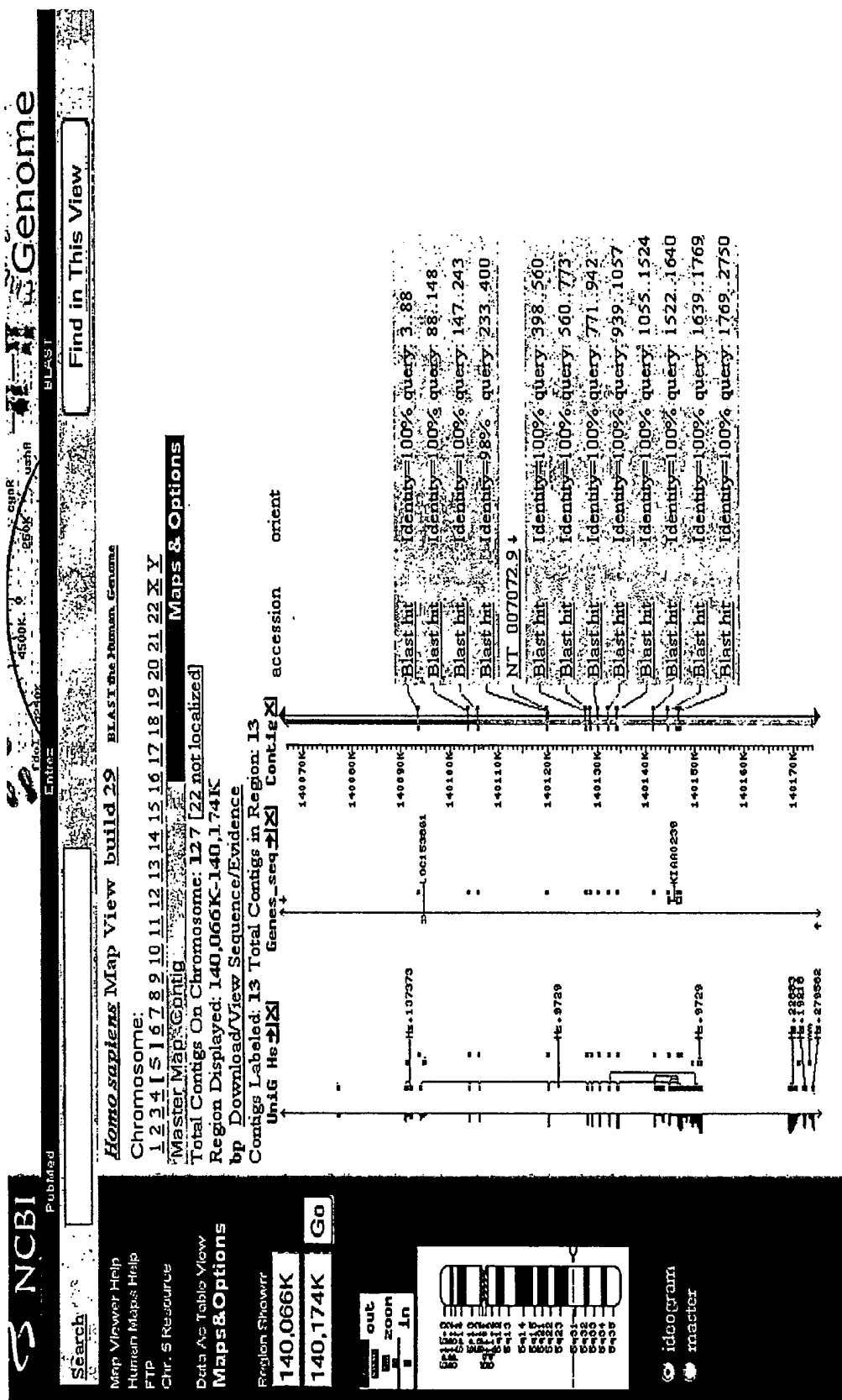


FIG. 6B







10/524426

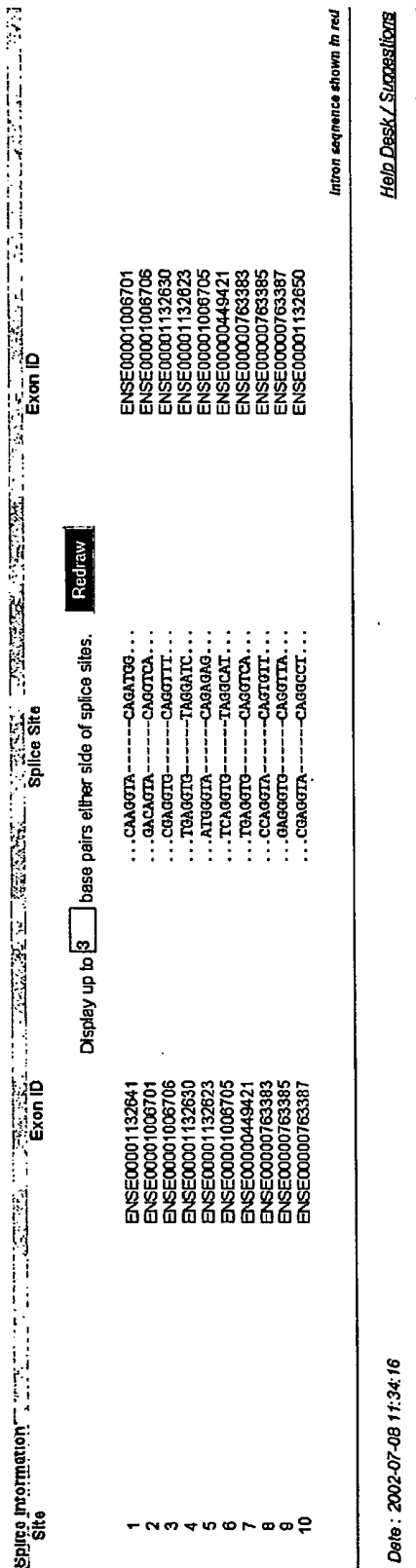



FIG. 6E

10/524426



Evidence Viewer

KIAA0239

PubMed

Nucleotide

Protein

OMIM

Genome

Taxonomy

PopSet

[Go to full display with alignments](#)

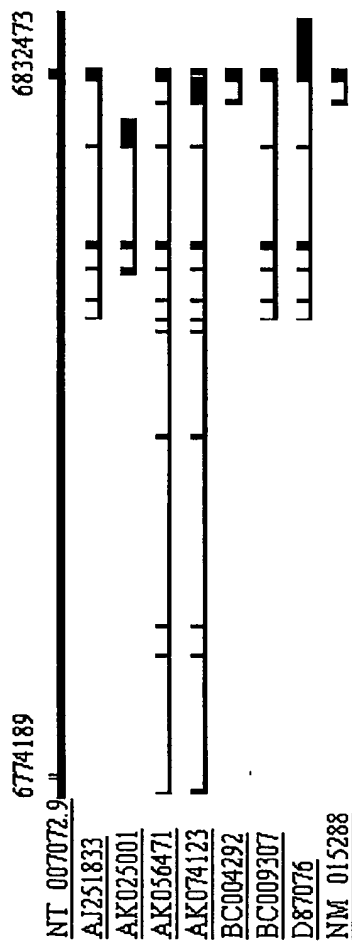


FIG. 6F

Simple  
Module  
Architecture  
Research  
Tool



SMART is a tool for the analysis of protein domain architecture. It is a part of the SMART database.

HOME LITERATURE ABOUT SMART WHAT'S NEW FEEDBACK

Currently there are too many SMART jobs running.

Your job is no. 1 in the queue.

Domains within the query sequence of 850 residues



Use the over domain / undefined region to see the history; click on it to go to further statistics; right-click to save whole protein as PNG image  
Transmembrane segments as predicted by the TMHMM program (TMH), coiled coil regions determined by the Coiled coil program (Coiled) and Signal peptides determined by the SignalP program (SignalP) are indicated by (---). Hits only found by BLAST are indicated by (---). Hits for this in the eukaryotic database and (---) for hits against PDB. Regions containing repeats detected by RepeatFinder but not covered by domains are indicated by (---).

## Architecture analysis

Display all proteins with similar domain organization.  
Display all proteins with similar domain composition.

The SMART diagrams above represents a summary of the results shown below. Domains with scores less significant than established cutoffs are not shown in the diagram. Features are also not shown when two or more occupy the same place of sequence; the priority for display is given by SMART > PFAM > PROSITE > Signal peptides > Transmembrane > Coiled coil > Low complexity. In either case, features not shown in the above diagram are marked 'Hidden'.

Confidently predicted domains, repeats, motifs and features

name	begin	end	E-value
low complexity	25	44	-
low complexity	177	193	-
PHD	217	263	1.35e-10
PHD	326	381	1.15e-04
low complexity	550	566	-
low complexity	661	671	-
low complexity	690	711	-
low complexity	727	741	-
low complexity	744	753	-

FIG. 6G

10/324426

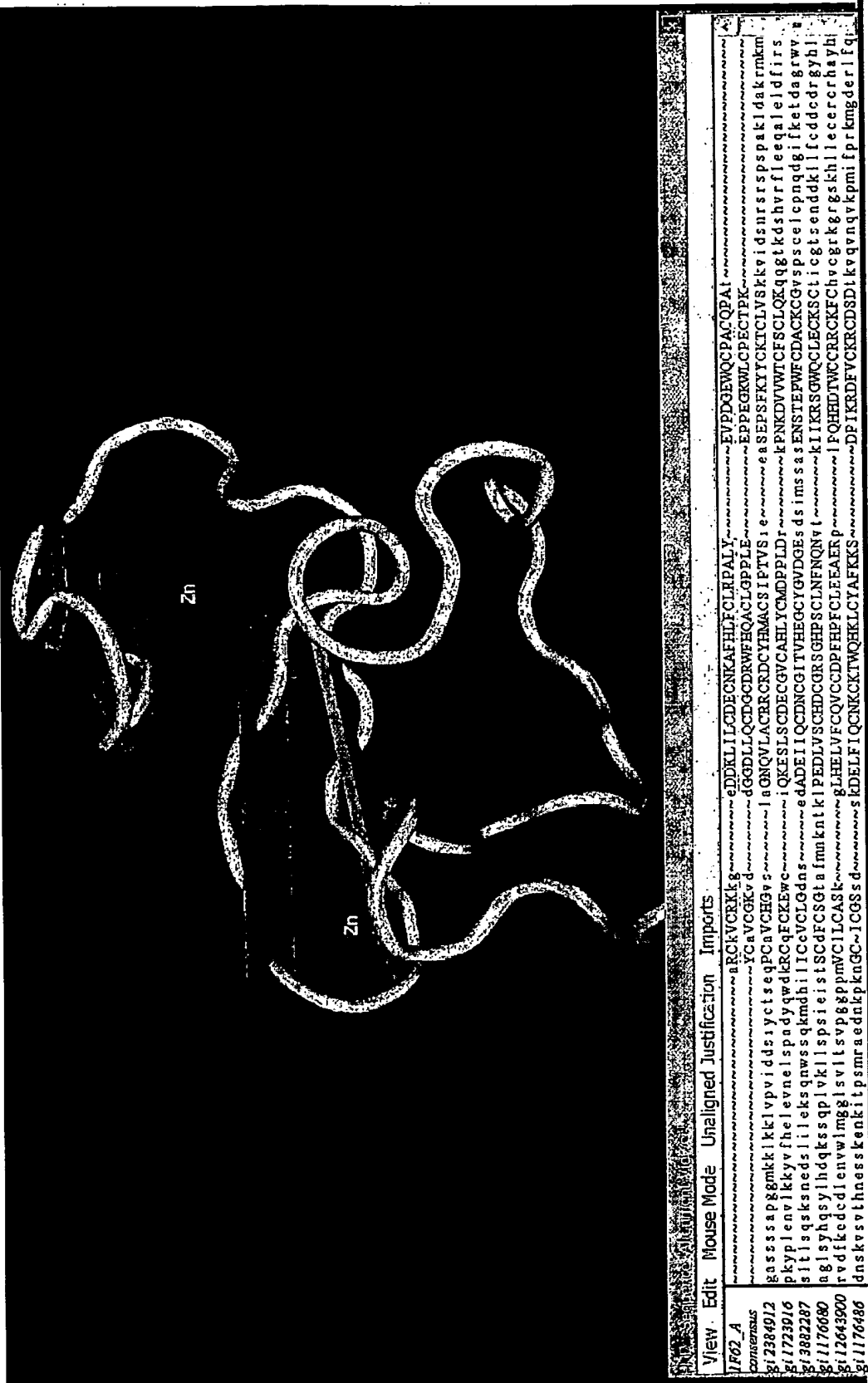


FIG. 6H

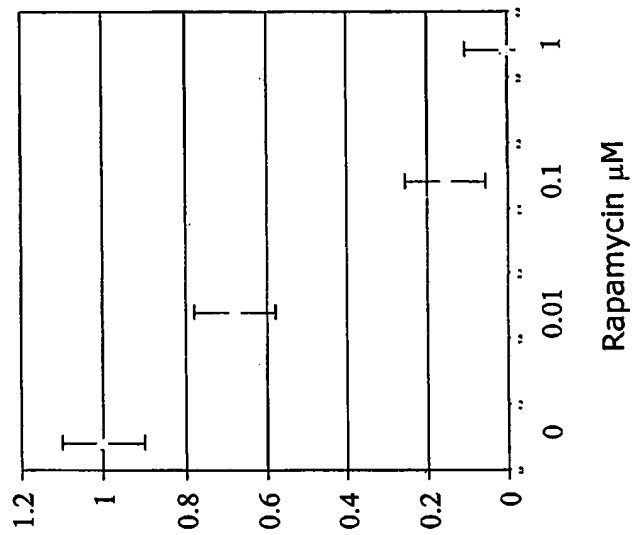


FIG. 7

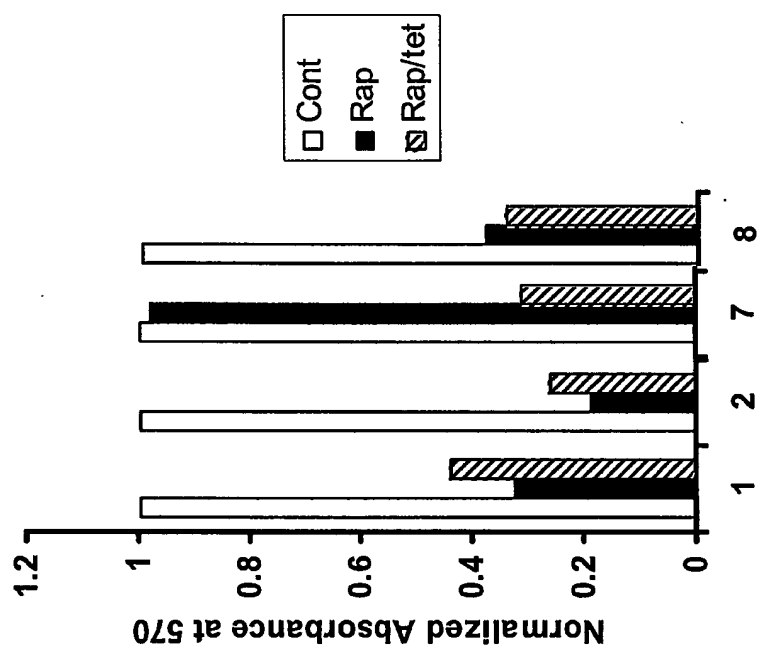


FIG. 8A



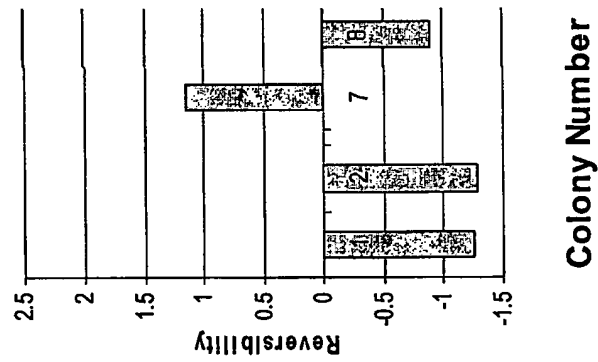


FIG. 8B

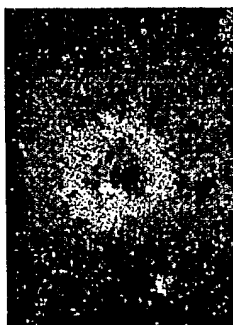


FIG. 8C

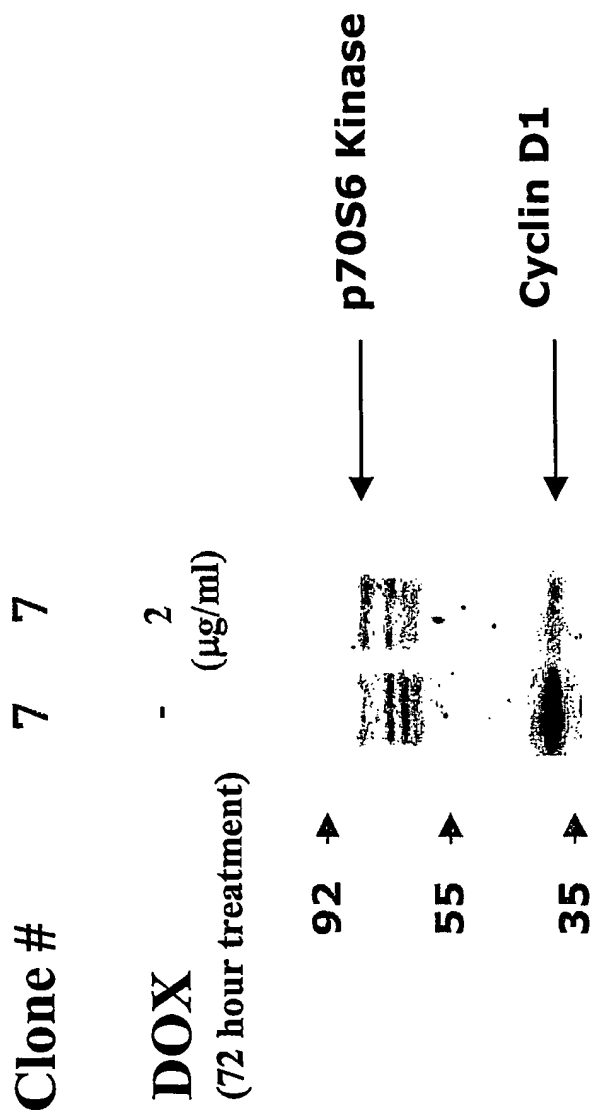


FIG. 8D

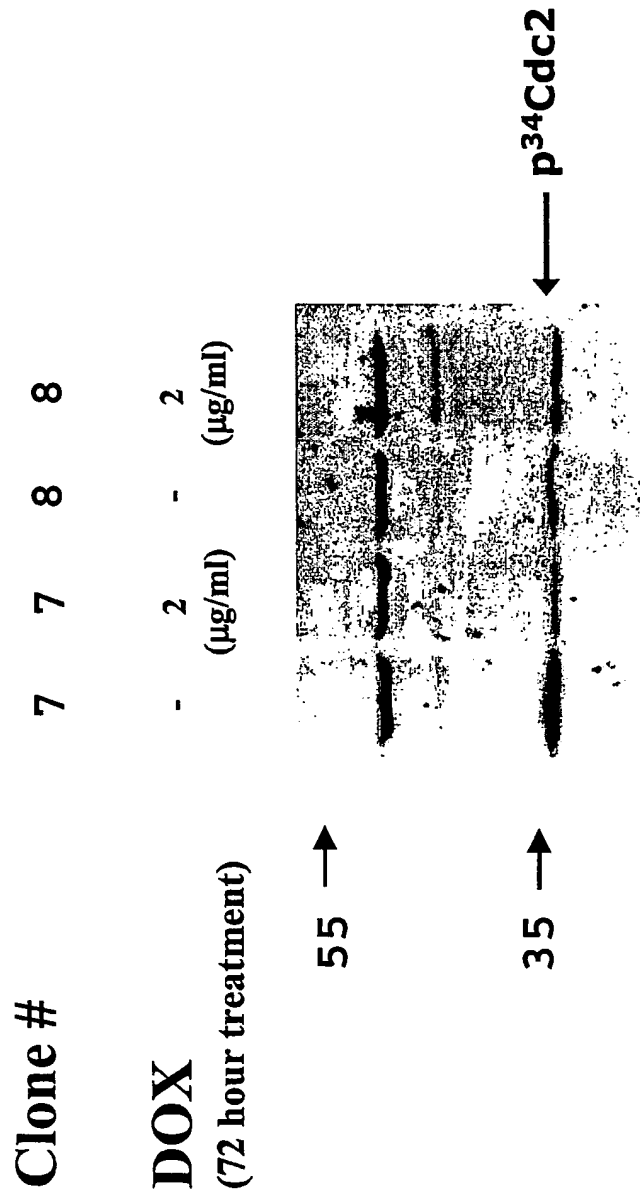


FIG. 8E



FIG. 9A

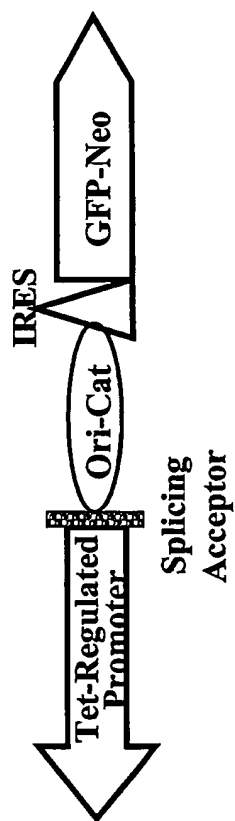


FIG. 9B

10/324,426

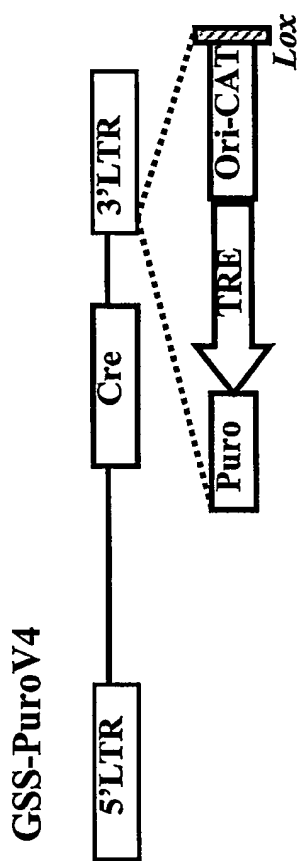


FIG. 9C

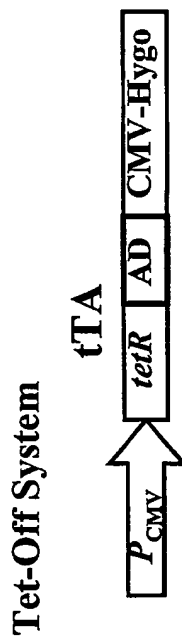


FIG. 9D

## SNP list:

rs167900  
rs179278  
rs191294  
rs329116  
rs329117  
rs329118  
rs329119  
rs329120  
rs329121  
rs329122  
rs329123  
rs329124  
rs329125  
rs329126  
rs329302  
rs329303  
rs329304  
rs329305  
rs329306  
rs329307  
rs329308  
rs329309  
rs329310  
rs329311  
rs329312  
rs329313  
rs329314  
rs329315  
rs329316  
rs329317  
rs329318  
rs329319  
rs329320  
rs329321  
rs329322  
rs329323  
rs329324  
rs329325  
rs329326  
rs329327  
rs592548  
rs620498  
rs885324  
rs1007342  
rs1044263  
rs1044264  
rs1966789  
rs2044318  
rs2084007  
rs2241699  
rs2241700  
rs2277063  
rs2304082  
rs2304083  
rs2304084  
rs2569340  
rs2569341  
rs2569343  
rs2589399  
rs2589400  
rs2589402  
rs2589403  
rs2589404  
rs2589405  
rs2589406  
rs2589408  
rs2589409

FIG. 10

## Alternative Splicing-1

### Overview

This gene is defined by 155 cDNA clones and 162 sequences. It is located on chromosome 5 on the direct strand, from base 140093473 to base 140150963. According to RefSeq annotation, its cytogenetic location is 5q31.2. The gene covers 57490 bp of genomic DNA. It produces, by alternative splicing, 12 different transcripts a, b, c, d, e, f, g, h, i, j, k, l, altogether encoding 11 proteins.

<http://www.ncbi.nlm.nih.gov/IEB/Research/Acembly/av.cgi?db=29&c=gene&a=fiche&l=G t5 Hs5 7229 29 18 2019 -#structure>

It contains 23 confirmed introns, 19 of which are alternative. Comparison to the genome sequence shows that 18 introns follow the consensual [gt-ag] rule, 1 the less frequent consensus [gc-ag], 2 are atypical with good support [ct\_gc], [ga\_ct] (provided there is no error in the genome), 2 are fuzzy or ill defined.

The gene gives rise to 12 types of transcripts, predicted to encode 11 distinct proteins. <http://www.ncbi.nlm.nih.gov/IEB/Research/Acembly/av.cgi?db=29&c=gene&a=fiche&l=G t5 Hs5 7229 29 18 2019 -#>

mRNA(s) and Protein(s)					
Transcript	5' complete	Sequence incomplete	3' complete	# exons	Transcr. unit bp
variant a	5'UTR=40bp		3'UTR=3435bp, polyA	15	57491
variant b	5'UTR=1162bp		3'UTR=158bp, polyA	12	53717
variant c	5'UTR=339bp		3'UTR=158bp, polyA	11	53717
variant d	5'UTR=339bp		3'UTR=158bp, polyA	10	53717
variant e	5'UTR=339bp		3'UTR=423bp, polyA	7	53717
variant f	5'UTR=346bp		3'UTR=1778bp, polyA	3	53717
variant g	5'UTR=339bp		3'UTR=3435bp, polyA	15	57473
variant h	5'UTR=339bp		3'UTR=5848bp, polyA	14	57473
variant i	5'UTR=339bp	2 exons inferred	3'UTR=3435bp, polyA	15	57473
variant j	5'UTR=339bp		3'UTR=3914bp, polyA	11	57473
variant k	no evidence		3'UTR=3914bp, polyA	11	57473
variant l	5'UTR=339bp	2 exons inferred	3'UTR=1113bp, polyA	12	53717
Protein	Starts on	Ends on	coord. on mRNA	specific clone(s)	
a complete	Met	Stop	41 to 2593	"AK056471.1"	
b complete	Met	Stop	1163 to 2839	"BM911233.1"	
c complete	Met	Stop	340 to 2712	"BC009307.1"	
d complete	Met	Stop	340 to 2460	2	
e complete	Met	Stop	340 to 927	"IMAGE:2162992"	
f complete	Met	Stop	347 to 1207	6	
g complete	Met	Stop	340 to 2844		
h complete	Met	Stop	340 to 2028	4	
i complete	Met	Stop	340 to 2841	2	
j complete	Met	Stop	340 to 2715		
k	1st codon	Stop	3 to 2564	"BM803449.1"	
l complete=h	Met	Stop	340 to 2028	2	

FIG. 11A



**Alternative Splicing-1****Intron exon structure and support**

	Length	Coord on gene	Supporting Clone(s)	Coord. on Clone
Alternative exon 1	88	1 to 88	<u>AK056471.1</u>	1 to 88
Alternative exon 2	339	19 to 357		
Alternative intron [gt-ag]	10115	89 to 10203		
Alternative intron [gt-ag]	9846	358 to 10203		
Alternative exon 3	188	445 to 632		
Alternative intron [gt-ag]	9571	633 to 10203		
Exon 4	58	10204 to 10261	<u>AK056471.1</u>	89 to 146
			<u>IMAGE:4541316</u>	339 to 396
			<u>IMAGE:5538609</u>	400 to 457
			<u>BM552145.1</u>	333 to 390
			<u>BM803449.1</u>	214 to 271
Intron [gt-ag]	2073	10262 to 12334		
Exon 5	95	12335 to 12429	<u>AK056471.1</u>	147 to 241
			<u>IMAGE:5538609</u>	458 to 552
			<u>BM552145.1</u>	391 to 485
			<u>BM803449.1</u>	272 to 366
			<u>IMAGE:4541316</u>	397 to 491
Intron [gt-ag]	13968	12430 to 26397		
Exon 6	158	26398 to 26555	<u>AK056471.1</u>	242 to 399
			<u>AW177999.1</u>	24 to 181
			<u>AW178006.1</u>	24 to 181
			<u>AW178035.1</u>	24 to 181
			<u>AW178043.1</u>	24 to 181
Intron [gt-ag]	7534	26556 to 34089		
Exon 7	161	34090 to 34250	<u>AK056471.1</u>	400 to 560
			<u>AK074123.1</u>	643 to 803
			<u>AW177999.1</u>	182 to 342
			<u>AW178043.1</u>	182 to 342
			<u>AW178053.1</u>	182 to 342
Intron [gt-ag]	755	34251 to 35005		
Alternative exon 8	90	35006 to 35095		
Alternative exon 9	110	35006 to 35115	<u>TCBAP0644</u>	157 to 266
Alternative exon 10	212	35006 to 35217	<u>AK056471.1</u>	561 to 772
			<u>AK074123.1</u>	804 to 1015
			<u>AW177999.1</u>	343 to 554
			<u>AW178035.1</u>	343 to 554
			<u>AW178043.1</u>	343 to 554
Alternative intron Fuzzy	18191	35096 to 53286		
Alternative intron [ga-ct]	5257	35116 to 40372		
Alternative intron [gt-ag]	1242	35218 to 36459		
Alternative exon 11	168	36460 to 36627	<u>AK056471.1</u>	773 to 940
			<u>AK074123.1</u>	1016 to 1183
			<u>AJ251833.1</u>	121 to 288
			<u>BC009307.1</u>	46 to 213
			<u>D87076.1</u>	26 to 193
Alternative intron [gt-ag]	2126	36628 to 38753		
Alternative intron [gt-ag]	2131	36628 to 38758		

FIG. 11B

WO 2004/020581

PCT/US2003/026073

34/35

Alternative exon 12	439	38432 to 38870	<u>AK025001.1</u>	1 to 439
Alternative exon 13	117	38754 to 38870	<u>AK056471.1</u>	941 to 1057
			<u>AK074123.1</u>	1184 to 1300
			<u>AJ251833.1</u>	289 to 405
			<u>BC009307.1</u>	214 to 330
			<u>D87076.1</u>	194 to 310
			<u>BM911233.1</u>	523 to 634
Alternative exon 14	112	38759 to 38870		
Alternative intron [gt-ag]	1505	38871 to 40375		
Alternative exon 15	468	40373 to 40840		
Alternative exon 16	465	40376 to 40840	<u>AK056471.1</u>	1058 to 1522
			<u>AK074123.1</u>	1301 to 1765
			<u>AJ251833.1</u>	406 to 870
			<u>AK025001.1</u>	440 to 904
			<u>BC009307.1</u>	331 to 795
Alternative intron [gt-ag]	7064	40841 to 47904		
Alternative exon 17	118	47905 to 48022	<u>AK056471.1</u>	1523 to 1640
			<u>AK074123.1</u>	1766 to 1883
			<u>AJ251833.1</u>	871 to 988
			<u>BC009307.1</u>	796 to 913
			<u>D87076.1</u>	776 to 893
Alternative exon 18	2081	47905 to 49985		
Alternative intron [gt-ag]	3005	48023 to 51027		
Alternative intron [gt-ag]	4731	48023 to 52753		
Alternative intron [gt-ag]	4734	48023 to 52756		
Alternative exon 20	129	51028 to 51156	<u>AK056471.1</u>	1641 to 1769
			<u>IMAGE:4806324</u>	517 to 646
			<u>BM853385.1</u>	88 to 216
			<u>IMAGE:4753761</u>	76 to 204
			<u>BF762782.1</u>	230 to 358
Alternative exon 21	271	51028 to 51298	<u>NM 015288.1</u>	7 to 277
			<u>BC004292.1</u>	7 to 277
Alternative exon 22	5215	51028 to 56242		
Alternative intron [gt-ag]	1597	51157 to 52753		
Alternative intron [gt-ag]	1600	51157 to 52756		
Alternative intron [gt-ag]	1458	51299 to 52756		
Alternative exon 23	982	52754 to 53735		
Alternative exon 24	3489	52754 to 56242		
Alternative exon 25	4738	52754 to 57491		
Alternative exon 26	979	52757 to 53735		
Alternative exon 27	3486	52757 to 56242		
Alternative exon 28	449	53287 to 53735	<u>IMAGE:2162992</u>	453 to 4
Alternative intron [ct-gc]	283	56243 to 56525		
Alternative exon 29	59	56526 to 56584	<u>IMAGE:4045343</u>	348 to 406
Alternative intron [gc-ag]	54	56585 to 56638		
Alternative exon 30	180	56639 to 56818	<u>IMAGE:4045343</u>	407 to 589
Alternative intron Puzzy	142	56819 to 56960		
Alternative exon 31	531	56961 to 57491		

FIG. 11B (Continued)

## PROTEIN ANALYSIS

<http://www.ncbi.nlm.nih.gov/IEB/Research/Acembly/av.cgi?db=29&c=mRNA&a=fic>

[he&l=G t5 Hs5 7229 29 18 2019.a - #](http://www.ncbi.nlm.nih.gov/IEB/Research/Acembly/av.cgi?db=29&c=mRNA&a=fic) Conceptual translation, MW, pI

The complete protein encoded between the first Met and the stop codon contains 850 residues. The calculated molecular weight of the protein is 93.6 kDa and isoelectric point 5.4.

<http://www.ncbi.nlm.nih.gov/IEB/Research/Acembly/av.cgi?db=29&c=mRNA&a=fic>

[he&l=G t5 Hs5 7229 29 18 2019.a - #](http://www.ncbi.nlm.nih.gov/IEB/Research/Acembly/av.cgi?db=29&c=mRNA&a=fic)

Predicted cellular localization and motifs (Psort)

PSORT II analysis, (K. Nakai <http://psort.nibb.ac.jp>) trained on yeast data and run on May 26, 2002, predicts that the subcellular location of this protein is most likely in the nucleus (69). Less likely possibilities are in the cytoplasm (17%) or in the mitochondria (4%) or in vesicles of secretory system (4%) or in the endoplasmic reticulum (4%). The following domains were found:

from aa to	domain	[sequence]
20	23 Nuclear_localization_domain	[KRRK]
163	190 Coil_coil_4	[ELINSELKEMERPELDELTLERVLEELE] (SEQ ID NO: 24)
514	522 2nd_peroximal_domain	[KLQEQIFHL]
548	564 Nuclear_localization_domain	[KRKGCEGSGSTEKKEK] (SEQ ID NO: 25)
549	565 Nuclear_localization_domain	[RKGCBSKSGSTEKKEKV] (SEQ ID NO: 26)
650	656 Nuclear_localization_domain	[PARKARG]
661	667 Nuclear_localization_domain	[PAKKKFP]
663	666 Nuclear_localization_domain	[KKKP]

<http://www.ncbi.nlm.nih.gov/IEB/Research/Acembly/av.cgi?db=29&c=mRNA&a=fic>

[he&l=G t5 Hs5 7229 29 18 2019.a - #](http://www.ncbi.nlm.nih.gov/IEB/Research/Acembly/av.cgi?db=29&c=mRNA&a=fic)

Protein family classification (Pfam)

Pfam analysis (<http://pfam.wustl.edu>) run on May 27, 2002, shows a significant hit to the PHD-finger from 217 to 265, with score 62.6 and E = 5.2e-16.

68 other expressed genes in the database also contain this motif

The PHD finger [MEDLINE:95216093] , [PUB00005675] is a C4HC3 zinc-finger-like motif found in nuclear proteins thought to be involved in chromatin-mediated transcriptional regulation. The PHD finger motif is reminiscent of, but distinct from the C3HC4 type RING finger. The function of this domain is not yet known but in analogy with the LIM domain it could be involved in protein-protein interaction and be important for the assembly or activity of multicomponent complexes involved in transcriptional activation or repression. In similarity to the RING finger and the LIM domain, the PHD finger is thought to bind two zinc ions.

[ 1 Trends Biochem Sci 1995;20:56-59.

]:

[ 2 J Mol Biol 2000;304:723-729.

]:

There are also 2 non significant Pfam hits.

<http://www.ncbi.nlm.nih.gov/IEB/Research/Acembly/av.cgi?db=29&c=mRNA&a=fic>

[he&l=G t5 Hs5 7229 29 18 2019.a - #](http://www.ncbi.nlm.nih.gov/IEB/Research/Acembly/av.cgi?db=29&c=mRNA&a=fic)

Protein homologies (BlastP)

BlastP analysis, run at NCBI on the non redundant database on May 27, 2002, shows 228 hits with expectancy less than 0.001. Interesting hints from this analysis are:

score	occurrences	
1001	1	FLJ00195 protein
600	1	FLJ22479
404	2	bromodomain

FIG.11C

**This Page is Inserted by IFW Indexing and Scanning  
Operations and is not part of the Official Record**

**BEST AVAILABLE IMAGES**

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- ☐ **BLACK BORDERS**
- ☐ **IMAGE CUT OFF AT TOP, BOTTOM OR SIDES**
- ☐ **FADED TEXT OR DRAWING**
- ☒ **BLURRED OR ILLEGIBLE TEXT OR DRAWING**
- ☐ **SKEWED/SLANTED IMAGES**
- ☐ **COLOR OR BLACK AND WHITE PHOTOGRAPHS**
- ☐ **GRAY SCALE DOCUMENTS**
- ☐ **LINES OR MARKS ON ORIGINAL DOCUMENT**
- ☐ **REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY**
- ☐ **OTHER:** \_\_\_\_\_

**IMAGES ARE BEST AVAILABLE COPY.**

**As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.**